Environment-to-phenotype mapping and adaptation strategies in varying environments

BingKan Xue[a,b,c,1], Pablo Sartori[a,b,c], and Stanislas Leibler[a,b,c,1]

[a]The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ 08540; [b]Laboratory of Living Matter, The Rockefeller University, New York, NY 10065; and [c]Center for Studies in Physics and Biology, The Rockefeller University, New York, NY 10065

Biological organisms exhibit diverse strategies for adapting to varying environments. For example, a population of organisms may express the same phenotype in all environments ("unvarying strategy") or follow environmental cues and express alternative phenotypes to match the environment ("tracking strategy"), or diversify into coexisting phenotypes to cope with environmental uncertainty ("bet-hedging strategy"). We introduce a general framework for studying how organisms respond to environmental variations, which models an adaptation strategy by an abstract mapping from environmental cues to phenotypic traits. Depending on the accuracy of environmental cues and the strength of natural selection, we find different adaptation strategies represented by mappings that maximize the long-term growth rate of a population. The previously studied strategies emerge as special cases of our model: The tracking strategy is favorable when environmental cues are accurate, whereas when cues are noisy, organisms can either use an unvarying strategy or, remarkably, use the uninformative cue as a source of randomness to bet hedge. Our model of the environment-to-phenotype mapping is based on a network with hidden units; the performance of the strategies is shown to rely on having a high-dimensional internal representation, which can even be random.

evolutionary theory | fluctuating environments | phenotypic plasticity | population dynamics | survival strategies

To study the properties of a physical system, a phenomenological approach is to characterize how it responds to external conditions. For instance, materials show particular patterns of deformation under external forces, which reveals their elastic properties. Biological organisms exhibit far more complex responses to environmental conditions. As the environment varies, organisms adapt by changing their phenotypes, including morphological and behavioral traits. Such phenotypic responses to the environment are modified through the process of evolution, which gives rise to different forms of adaptation. Several adaptation strategies, as described below, have been studied both experimentally and theoretically (1–8). In this paper, we adopt the phenomenological approach to study biological adaptation by modeling general forms of phenotypic responses to environmental conditions. This approach enables us to reveal underlying connections between different adaptation strategies.

The simplest adaptation strategy is one in which organisms express the same phenotype in all environments. A population using this strategy has a narrow distribution of phenotypes that does not vary with the environment. In such an "unvarying strategy," the typical phenotype is often fit for most environmental conditions. For example, birds that feed on a variety of food sources ("generalists") have a midsized beak, which is slender enough for catching insects and conical enough for cracking seeds (9, 10).

Another strategy is for organisms to follow environmental cues and express alternative phenotypes to match the environment. Provided that the cues are accurate, individual organisms of a population may all express the appropriate phenotype. The phenotype distribution would thus exhibit a narrow peak that tracks the environmental variation. Examples of this "tracking strategy" are seasonal changes of the butterfly's wing patterns and the mammal's coat colors, which are induced by weather conditions and provide suitable camouflage (11).

A third strategy is such that individual organisms of the same population express different phenotypes, so that the phenotype distribution is broad or has multiple peaks. Such diversification is useful in stochastically changing environments, since there will always be some individuals in the population that have the right phenotype to survive. A classical example of this "bet-hedging strategy" is the seed bank: To cope with unpredictable inclement weather, some seeds quickly germinate after being dispersed while others remain in the soil for a prolonged period (12, 13). Bet hedging can also be combined with cue tracking, such that the distribution of phenotypes varies according to the environment. For example, the fraction of seeds that germinate can depend on environmental factors such as temperature, moisture, and the presence of other seeds (14, 15).

We show that the above strategies are special limits of a general solution for adaptation to varying environments. Depending on the accuracy of environmental cues and the strength of natural selection, particular strategies of adaptation emerge from a continuum of possible strategies. This unifying picture is obtained using a model of "environment-to-phenotype mapping," which allowed us to explore a wide range of phenotypic responses to environmental conditions. Essential to our model is a high-dimensional internal representation of the environment that allows organisms to develop diverse phenotypic responses.

Significance

A fundamental difference between living and nonliving systems is that organisms can evolve responsive adaptation to external conditions. We present a theoretical framework, which unifies different adaptation strategies encountered in biology. Central to our approach is the introduction of an environment-to-phenotype mapping describing how organisms' traits or behavior depend on the environment. In contrast to commonly considered genotype-to-phenotype mapping, our approach emphasizes an evolutionary rather than mechanistic understanding of organisms. Our phenomenological model, inspired by artificial neural networks, also allows us to study the importance of the dimensionality of internal representations for the adaptation strategies.

Our results suggest ways to experimentally evolve and identify different adaptation strategies.

Model of Environment-to-Phenotype Mapping

The phenotypic responses of an organism to environmental conditions can be conceptualized as a mapping from the environment space to the phenotype space. A certain environmental stimulus that the organism experiences may induce a particular phenotype. Such an environment-to-phenotype mapping may represent, for example, how the development of organisms is affected by the environment (known as "phenotypic plasticity"). A mapping that allows a population to survive better and reach greater abundance in the long term will generally be favored by natural selection. We study the optimal form of the mapping that maximizes the population growth rate in varying environments.

Consider a population of organisms that reproduce asexually in discrete numbers of generations. The environment they live in may vary from generation to generation. An environmental condition is described by an n-dimensional vector ε, whose components represent different environmental factors, such as temperature, light, and amount of food. We assume that the environment switches between several different conditions, labeled by ε^μ for $\mu = 1, \ldots, m$. Each individual organism receives an environmental cue, which is correlated with the environmental condition and can potentially be used to distinguish the actual environment. This environmental cue is denoted by a vector ξ, which is assumed to belong to the n-dimensional environment space. Note that, in the same environment ε^μ, each organism may receive a different cue ξ.

Similarly, the phenotype of an organism is described by a p-dimensional vector ϕ, whose components represent different characteristic traits, such as the shape of body parts or the speed of movement. The phenotype that an organism expresses may depend on the environmental cue ξ that it receives. We describe such dependence by a function, $\phi = \Phi(\xi)$, which represents a mapping from the n-dimensional environment space to the p-dimensional phenotype space, as illustrated in Fig. 1A. Different forms of the mapping will correspond to different adaptation strategies.

The fitness of an organism in a given environment ε^μ is measured by how many offspring it produces. This depends on its phenotype ϕ and is described by a function $f(\phi; \varepsilon^\mu)$. Thus, in each generation, labeled by a number t, an individual organism that receives an environmental cue ξ_t will express a phenotype $\phi_t = \Phi(\xi_t)$ and produce as many as $f(\Phi(\xi_t); \varepsilon_t)$ offspring, where ε_t is the environmental condition. Let N_t be the population size in the tth generation; then in the next generation it will be

$$N_{t+1} = N_t \sum_{\xi_t} P(\xi_t \mid \varepsilon_t) f(\Phi(\xi_t); \varepsilon_t), \qquad [1]$$

where $P(\xi_t \mid \varepsilon_t)$ is the probability that a cue ξ_t is received when the environment is ε_t. In the long term, the growth rate of the population is given by $\Lambda \equiv \frac{1}{T} \log \frac{N_T}{N_0}$ for $T \to \infty$. This long-term growth rate can be calculated as

$$\Lambda = \sum_\mu p_\mu \log \sum_\xi P(\xi \mid \varepsilon^\mu) f(\Phi(\xi); \varepsilon^\mu), \qquad [2]$$

where p_μ is the probability that each environmental condition ε^μ occurs. We use Λ as the measure of evolutionary success for a population. The optimal phenotypic response is determined by the function Φ that maximizes the value of Λ.

For simplicity, we assume that the environmental cue ξ is randomly distributed around the actual environment ε^μ according to a Gaussian distribution, $P(\xi \mid \varepsilon^\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{(\xi - \varepsilon^\mu)^2}{2\sigma^2}\}$, where σ represents the noisiness of the environmental cue. The fitness is also assumed to be a Gaussian function, $f(\phi; \varepsilon^\mu) = F_\mu \exp\{-\frac{\gamma^2(\phi - \psi^\mu)^2}{2}\}$, where F_μ is a constant representing the maximum number of offspring in the environment ε^μ, and ψ^μ is
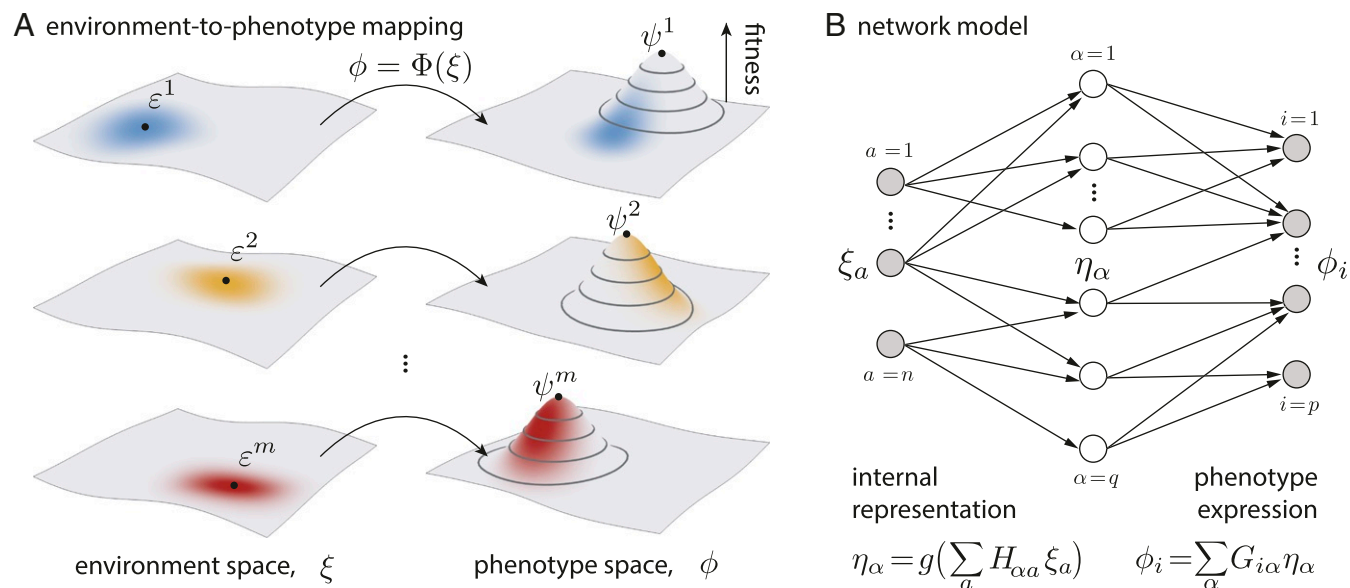

Fig. 1. Schematic illustration of our modeling framework. (A) Phenotypic responses described by a mapping from an n-dimensional environment space to a p-dimensional phenotype space. The environment can be in one of m conditions, labeled by ε^μ, each favoring a phenotype ψ^μ. In a given environmental condition (distinguished by color), each individual organism receives a noisy cue ξ (distribution represented by color shade in environment space) and expresses a phenotype according to the mapping Φ (distribution of phenotypes induced by the mapping is represented by color shade in phenotype space). The fitness of a phenotype depends on its distance to the favorable phenotype (illustrated by the fitness landscape in phenotype space). Note that the organism sees only the cue and does not know the true environment. (B) A network model with one hidden layer. The input ξ has n components, ξ_a. The hidden layer has q components, given by $\eta_\alpha = g(\sum_a H_{\alpha a} \xi_a)$, where $H_{\alpha a}$ is the representation matrix and g is a sigmoid function. The output ϕ has p components, determined by $\phi_i = \sum_\alpha G_{i\alpha} \eta_\alpha$, where $G_{i\alpha}$ is the expression matrix.

the most favorable phenotype in that environment. The parameter γ represents the strength of natural selection, which is assumed to be the same for all environments (see *SI Appendix,* *Fig. S3* for a different case). Note that σ and $1/\gamma$ serve as characteristic scales for the environment space and the phenotype space, respectively. Under those assumptions, the long-term growth rate Λ is evaluated numerically according to Eq. **2**, as described in *Materials and Methods*.

We are interested in the ideal function Φ^* that maximizes Λ, which satisfies the variational equation $\delta\Lambda/\delta\Phi(\xi) = 0$. Unfortunately, this equation cannot be solved explicitly in general (but see *Materials and Methods* for special cases). To proceed further, we need to specify the function Φ in a parametric form, so that we can optimize over the parameters numerically. The form of the function should be sufficiently general to allow all possible types of phenotypic responses. In the following, we introduce a particular form of the function that is biologically motivated as well as computationally convenient.

Our model of the function Φ takes the form of a feed-forward network with a hidden layer. The input layer has n nodes, corresponding to the n components of the environmental cue ξ; the output layer has p nodes, corresponding to the p components of the phenotype ϕ; the hidden layer is chosen to have q nodes, a potentially large number compared with n and p, as illustrated in Fig. 1*B*. These hidden nodes can be thought to form an internal representation of the external environment; their values are determined by the input vector ξ through a "representation matrix" H and a nonlinear transformation g, such as a tanh function. The output vector ϕ depends on the internal variables through an "expression matrix" G. Altogether, the function Φ takes the form

$$\phi_i = \Phi_i(\xi) = \sum_\alpha G_{i\alpha}\, g\left(\sum_a H_{\alpha a}\xi_a\right). \qquad [3]$$

(Each matrix has an additional column that represents a constant ["bias"] term; e.g., $\sum_a H_{\alpha a}\xi_a \equiv \sum_{a=1}^n H_{\alpha a}\xi_a + H_{\alpha 0}$, where $H_{\alpha 0}$ is the constant term that is optimized as part of the matrix.) With sufficiently many internal variables, such a multilayered feed-forward network [known as a "perceptron" (16)] can approximate any smooth function and hence capture all possible phenotypic responses.

The structure of this model is inspired by many biological systems. The hidden nodes of the network may represent internal variables of the organism. For example, a plant's phenotypic responses to environmental conditions can be described by a growth-regulatory network, where a large group of molecules, such as growth factors and gene promoters, act as hidden nodes of the network (17). The formation of a high-dimensional internal representation, which allows organisms to better perceive the environment and produce more refined phenotypic responses, has also been suggested. Cellular signaling networks, for example, involve many proteins that often have multiple modification sites, interacting with each other and giving rise to a large number of possible states (18). Similarly, biological neural networks, such as the olfactory systems of insects and mammals, have multiple layers of neurons for processing sensory information; some intermediate layers of neurons may play the role of expanding the dimensionality of input signals to facilitate later stages of cognition (19, 20).

In our network model, the environment-to-phenotype mapping is specified by the representation matrix H and the expression matrix G. These matrices may represent information that is encoded in the organism's genotype, which undergoes evolution. For simplicity, we consider the case where individuals of the population share the same matrices, and we look for the optimal values of H and G that maximize the long-term population growth rate Λ.

Emergence of Different Adaptation Strategies

The adaptation strategy resulting from the optimized network will depend on the level of environmental noise σ and the strength of natural selection γ. We explore the range of adaptation strategies in the (σ, γ) parameter space using numerical examples. Consider a 2D environment space ($n = 2$), a 3D phenotype space ($p = 3$), and a 20D internal space ($q = 20$). The environment switches between three conditions ($m = 3$), with arbitrarily chosen positions in the environment space (marked in Fig. 2*A*) and probabilities of occurrence ($p_\mu = 0.2$, 0.5, and 0.3, respectively). For each environmental condition ε^μ, we assign a most favorable phenotype ψ^μ, called an "archetype" hereafter, in the phenotype space (Fig. 2*B*). In a given environment ε^μ, organisms receive a distribution of cues, as illustrated in Fig. 2*A*. The mapping given by the optimized network generates a distribution of phenotypes, as illustrated in Fig. 2*B*. The shape of the phenotype distribution, and how it changes under different environmental conditions, characterizes the corresponding adaptation strategy.
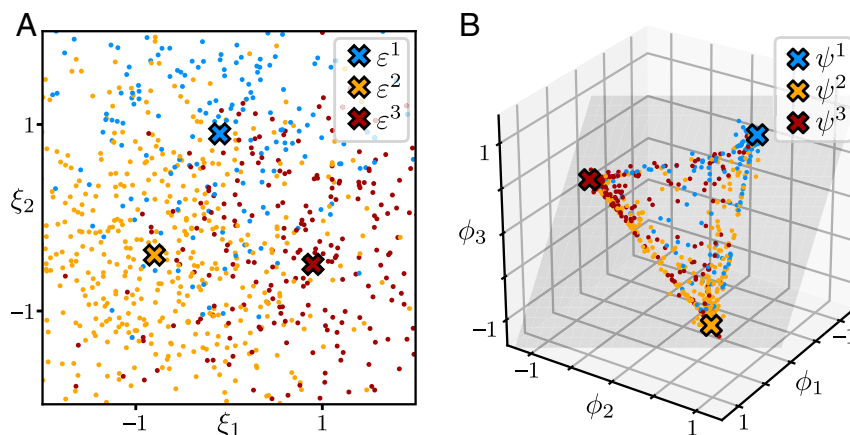

Fig. 2. Example of an adaptation strategy produced by an optimized network. (*A*) Distribution of environmental cues ξ represented by points in the environment space (color represents the actual environmental condition ε^μ). (*B*) Distribution of phenotypes produced by the optimized network, represented by points in the phenotype space. All points fall on a plane (gray shaded) spanned by the archetypes ψ^μ. For these plots we used parameter values $\sigma = 1$ for environmental noise and $\gamma = 1$ for selection strength, which represent characteristic scales that are of the same order as the distance between two environments ε^μ and between two archetypes ψ^μ, respectively.

A prominent feature of the emerged geometric structure, shown in Fig. 2B, is that all phenotypes lie on a flat plane spanned by the archetypes, $\{\psi^\mu\}$. This structure can be explained by a "Pareto efficiency" argument as follows. Since the fitness of a phenotype depends on its distance to the archetypes, a phenotype located off the plane will always be less fit than its perpendicular projection onto the plane. Therefore, in the optimal phenotype distribution, all phenotypes should fall on the plane. In general, if there are m archetypes, the optimal phenotype distribution will be contained in a $(m-1)$-dimensional subspace spanned by those archetypes. If m is small compared with the dimensionality of the original phenotype space, p, then the dimensionality of phenotypes will be significantly reduced.

Such dimensional reduction, as well as the Pareto efficiency argument, is similar to that found in the model in ref. 10. In that model, the archetypes represent different biological tasks that every individual organism must perform during its lifetime, with varied degrees of importance to its overall fitness. To compare with our model, we can associate the tasks with environmental conditions that individuals may encounter and need to adapt to, with varied probabilities of occurrence. From this perspective, the model in ref. 10 corresponds to the situation where the phenotype does not depend on the present environment (i.e., no phenotypic plasticity), and the phenotype distribution of a population is simply localized at a given point in the phenotype space. This form of phenotypic response and the resulting phenotype distribution are characteristic of the unvarying strategy, which is discussed later. In contrast, by allowing the phenotype to depend on environmental cues through the environment-to-phenotype mapping, our model encompasses a wider range of adaptation strategies, as we describe below.

Examples of Strategies. In the following, we examine the distribution of phenotypes for different parameters σ and γ, represented by the density of points in the archetype plane, as shown in Fig. 3 (see *SI Appendix*, Fig. S1 for clarity). In many cases, the density is high near the archetypes. We divide the plane into regions surrounding each ψ^μ, marked by boundary lines in Fig. 3; *Insets* show fraction of phenotypes lying inside each region. By comparing those fractions as well as the shape of the phenotype distribution between different environmental conditions, we identify a wide range of adaptation strategies.

Tracking strategy under low noise. Examples of low environmental noise are shown in Fig. 3 *G–I*. In these cases, the width of the noise distribution is much smaller than the typical distance between two environmental conditions (chosen to be $\simeq 1$); i.e., $\sigma \ll 1$. Therefore, the environmental cue is very accurate about the present environmental condition. As a result, in each environment ε^μ, the phenotype distribution is highly concentrated near the corresponding archetype ψ^μ—the surrounding region contains almost 100% of the phenotypes, so the plots in Fig. 3 *G–I*, *Insets* look diagonal. This means that the organisms can express the most favorable phenotype that tracks the varying environmental condition. The picture hardly changes as the selection strength γ is varied (compare among Fig. 3 *G–I*). It is understandable since, without a significant cost for sensing, organisms should always use environmental cues when those are reliable.

Unvarying strategy under high noise and weak selection. The opposite case where the environmental noise level is high ($\sigma \gg 1$) is shown in Fig. 3 *A–C*. In these examples, the environmental cue has a broad distribution and is largely uninformative about the actual environment. Therefore, we expect the optimal phenotype distributions to look similar in all environments. This is verified by Fig. 3 *A–C*, *Insets*, which show that there is a significant fraction of phenotypes in each region and the fractions vary slightly between different environments (see also *SI Appendix,*

Fig. S1). However, depending on the selection strength γ, the phenotype distribution has very different characters. Fig. 3A shows the case of weak selection, where the characteristic scale $1/\gamma$ is much larger than the typical distance between two phenotypes (chosen to be $\simeq 1$); i.e., $\gamma \ll 1$. In this case, the phenotypes are centered near the average phenotype, $\bar{\psi} = \sum_\mu p_\mu \psi^\mu$, regardless of the environmental condition. It means that the organisms may ignore the cue when it is noisy and exhibit a constant phenotype. The optimal constant phenotype strikes a balance between all of the archetypes, similar to the result in ref. 10.

Bet-hedging strategy under high noise and strong selection. When the cue is noisy and the selection is strong ($\sigma, \gamma \gg 1$), however, the unvarying strategy fails because the average phenotype $\bar{\psi}$ suffers from low fitness values in all environments. In this case, surprisingly, the organisms do not ignore the uninformative environmental cue, but use it in a completely different way—each organism expresses one of the archetypes according to the cue, so that the population diversifies into multiple subpopulations due to the randomness of the cue. As shown in Fig. 3C, the phenotype distribution is sharply peaked around every archetype ψ^μ, and the size of each peak changes little with the environmental condition. This bet-hedging strategy guarantees that, in any environment ε^μ, a subpopulation expressing the corresponding archetype ψ^μ will have a high fitness value. The relative size of each subpopulation depends on the probability p_μ that each environment occurs. In the limit of extremely strong selection ($\gamma \to \infty$), we expect to recover the result of previous bet-hedging models (e.g., ref. 21), in which the probability of expressing the archetype ψ^μ matches the probability of encountering the environment ε^μ (*Materials and Methods*). This is indeed the case, as the fraction of phenotypes near each ψ^μ agrees well with the environment probability p_μ (*SI Appendix,* Fig. S1C).

Intermediate strategies. Besides the above extreme cases that correspond to well-categorized adaptation strategies, intermediate cases are also found. A combination of bet-hedging and tracking strategies is seen in the case of a medium noise level ($\sigma \simeq 1$) and strong selection ($\gamma \gg 1$). As shown in Fig. 3F, the phenotype distribution is peaked around the archetypes, but the relative sizes of the peaks are biased toward the one that matches the actual environment (see also *SI Appendix,* Fig. S1F). This case may represent the situation of bet hedging with partial environmental information, in which the population uses an imperfect cue to moderately adjust its phenotype distribution (21, 22). Similarly, we can see intermediate cases between bet-hedging and unvarying strategies (high noise $\sigma \gg 1$ and medium selection $\gamma \simeq 1$, Fig. 3B), as well as between unvarying and tracking strategies (medium noise $\sigma \simeq 1$ and weak selection $\gamma \ll 1$, Fig. 3D).

The transition of adaptation strategies with the parameters σ and γ, illustrated by the examples in Fig. 3, can also be understood analytically using approximate solutions of the ideal function Φ^* for extreme parameter values (*Materials and Methods*). Those approximate solutions do not rely on the parametric form of the function, Eq. **3**, showing that our results are more general than the numerical examples. Generally, the accuracy of environmental cues, measured by the noise level σ, determines the bias of the phenotype distribution toward the archetype in a given environmental condition. The selection strength γ, on the other hand, modifies the shape of the phenotype distribution, which tends to be more clustered near the archetypes when the selection is strong and more scattered into the interior space between the archetypes when the selection is weak.

Quantification of Strategies. The shape of the phenotype distributions illustrated above can be characterized quantitatively.
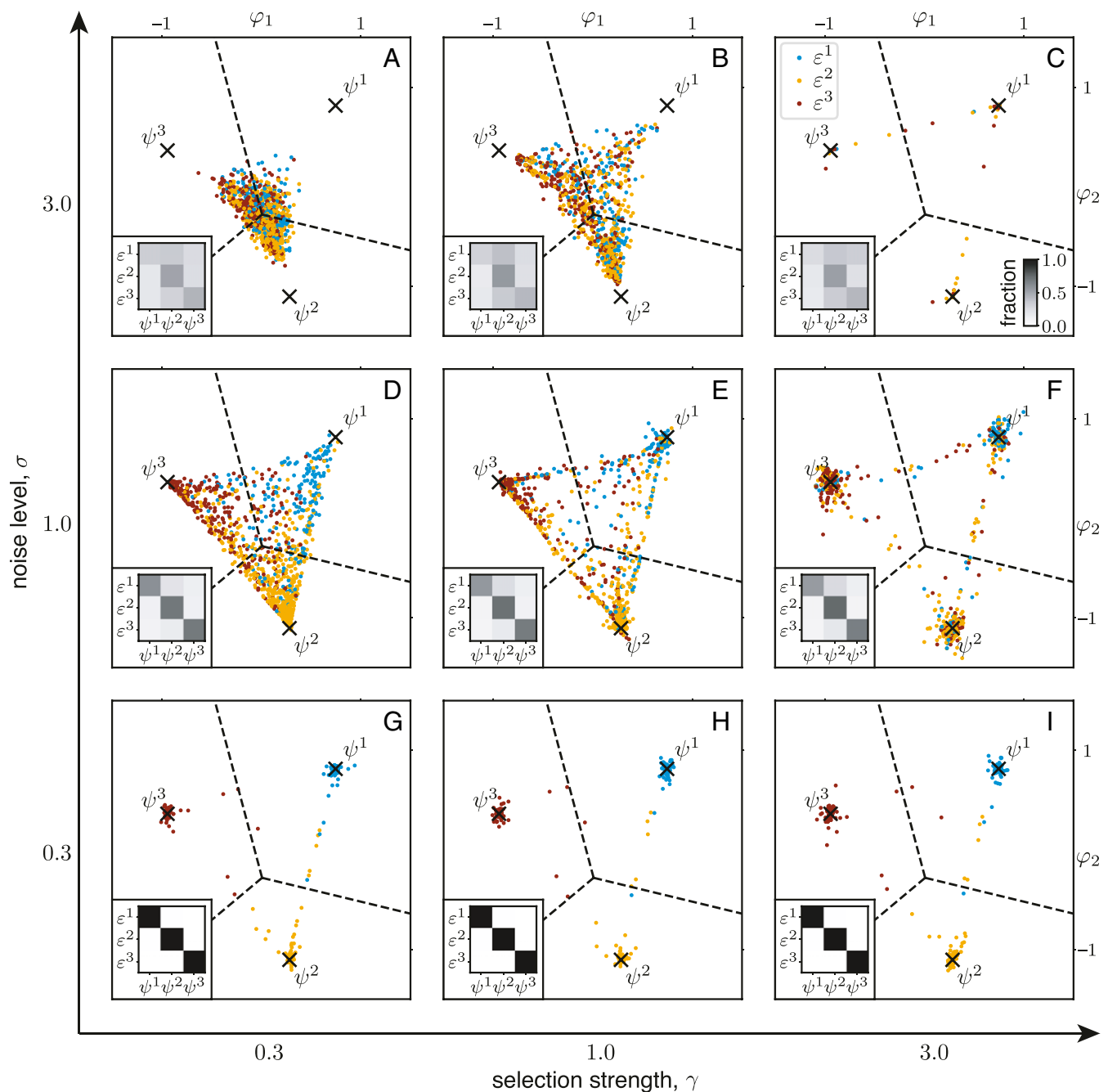
Fig. 3. (*A–I*) Phenotype distributions produced by networks optimized for different values of the noise level σ and the selection strength γ. Colored circles represent phenotypes plotted in the archetype plane, with φ_1, φ_2 as new coordinates. Dashed lines divide the plane into regions that are close to each ψ^μ; the intersection point corresponds to the average phenotype, $\bar\psi = \sum_\mu p_\mu \psi^\mu$. *Insets* show the fraction of phenotypes inside each region under different environmental conditions.

Two main properties of the phenotype distributions are how much they vary with the environment and how concentrated they are near the archetypes. To describe these properties, we introduce two characteristic quantities and examine how they vary with the environmental noise σ and the selection strength γ.

Specifically, in each environment ε^μ, the phenotype distribution can be denoted by a conditional probability distribution $\pi(\phi \,|\, \varepsilon^\mu)$, as defined in Eq. **15** (*Materials and Methods*). Given the environment probabilities p_μ, the overall distribution of the phenotype is $\pi(\phi) = \sum_\mu p_\mu \pi(\phi \,|\, \varepsilon^\mu)$. The total variance of the phenotype can be decomposed as $\mathbb{V}[\phi] = \mathbb{V}[\mathbb{E}[\phi \,|\, \varepsilon^\mu]] + \mathbb{E}[\mathbb{V}[\phi \,|\, \varepsilon^\mu]]$.

In the first term, $\mathbb{E}[\phi \,|\, \varepsilon^\mu]$ is the conditional expectation of the phenotype for a given environment ε^μ, and $\mathbb{V}[\mathbb{E}[\phi \,|\, \varepsilon^\mu]]$ is the variance of the conditional expectation with respect to the environment probabilities p_μ, and similarly for the second term. We can use these two terms to characterize different adaptation strategies. Essentially, the first term characterizes how much the phenotype varies with the environment, whereas the second term characterizes how much the phenotype varies in a given environment. For clarity, we take the trace of the variance matrices and normalize the terms by the variance of the archetypes, $\mathbb{V}[\psi]$ (according to the Pareto efficiency argument, the optimal phenotype distributions are contained in between
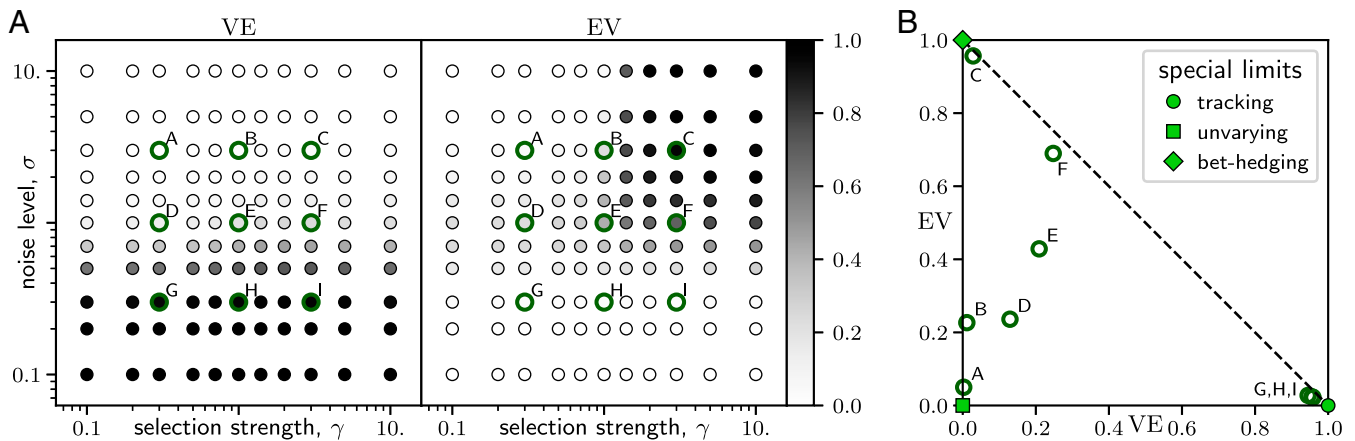
Fig. 4. Characterization of adaptation strategies using quantities VE and EV. (*A*) Plot of the parameter space showing how VE and EV vary with the noise level σ and the selection strength γ. Circles represent data points from numerical calculations, with values of VE and EV illustrated by grayscale; thick colored circles correspond to examples shown in Fig. 3. (*B*) Plot of VE−EV space showing examples from Fig. 3. Dashed line represents the bound VE + EV ≤ 1. Corners of the VE−EV space represent special limits that correspond to the tracking, unvarying, and bet-hedging strategies.

the archetypes; hence $\mathbb{V}[\phi] \leq \mathbb{V}[\psi]$). Thus, our characteristic quantities are

$$\mathrm{VE} \equiv \frac{\mathrm{tr}\left(\mathbb{V}\left[\mathbb{E}[\phi \mid \varepsilon^\mu]\right]\right)}{\mathrm{tr}\left(\mathbb{V}[\psi]\right)}, \quad \mathrm{EV} \equiv \frac{\mathrm{tr}\left(\mathbb{E}\left[\mathbb{V}[\phi \mid \varepsilon^\mu]\right]\right)}{\mathrm{tr}\left(\mathbb{V}[\psi]\right)}. \quad [4]$$

Fig. 4*A* shows how the values of these quantities change according to the parameters σ and γ.

To see how these quantities help characterize different adaptation strategies, consider the three strategies described above. For the tracking strategy, the phenotypes are concentrated near the corresponding archetype in each environment, and hence $\mathbb{E}[\phi \mid \varepsilon^\mu] \approx \psi^\mu$ and $\mathbb{V}[\phi \mid \varepsilon^\mu] \approx 0$; therefore, VE ≈ 1 and EV ≈ 0. Similarly, for the unvarying strategy, the phenotypes are always concentrated near the center of the archetypes, which means $\mathbb{E}[\phi \mid \varepsilon^\mu] \approx \bar{\psi}$ and $\mathbb{V}[\phi \mid \varepsilon^\mu] \approx 0$; therefore, VE ≈ 0 and EV ≈ 0. Finally, for the bet-hedging strategy, the phenotype distributions are largely independent of the environment and are concentrated near the archetypes in proportion to the environment probabilities p_μ; this leads to VE ≈ 0 and EV ≈ 1. Therefore, those three strategies can be clearly distinguished by different limits of the characteristic quantities, as shown in Fig. 4*B*.

Dimensionality of Internal Representation

So far we have fixed the dimensionality of the network's hidden layer at a relatively large number, $q = 20$, compared with that of the environment space, $n = 2$. The motivation was to create an adequate expansion of dimensionality from the input layer to the hidden layer, $q/n = 10$, so that the network can be used to approximate well the ideal function Φ^* in all cases. The approximation is verified in the limit $\gamma \to 0$, where explicit solutions can be found (*Materials and Methods*); the numerical solutions we obtained are very close to the ideal function Φ^* (*SI Appendix*, Fig. S2 *B* and *C*).

Let us now explore how the results change if we vary the dimensionality q. Fig. 5 shows how the maximum value of Λ increases with q. For a small q, the network model becomes very restrictive because it does not have many parameters that can be tuned. In that case, the phenotype distribution that results from optimizing the network will be deformed from that for the ideal function Φ^* (*SI Appendix*, Fig. S2*A*). In particular, in the limit $q \to 0$, the intermediate layer of the network vanishes, so the output becomes disconnected from the input. This means that the phenotype can no longer depend on the environmental cue, and hence the organism is forced to express the same phenotype in

all environments. In other words, the organism can use only the unvarying strategy, even though it is not favorable in many situations. On the other hand, a large q enables organisms to form various types of adaptation strategies, as we have seen for $q = 20$. The price, however, is having to tune a lot of parameters. This could mean a much longer time for a population to adapt to a varying environment.

In our numerical computation, we found that it is much slower to optimize over the representation matrix H than over the expression matrix G, because the latter is directly connected to the output phenotype being selected but the former is not. This suggests that it is harder for an organism to adjust the way it creates an internal representation of the environment than to adjust the mechanism that produces the phenotype directly. It is therefore interesting to ask whether one can keep the representation matrix H fixed while optimizing over the expression matrix G alone.
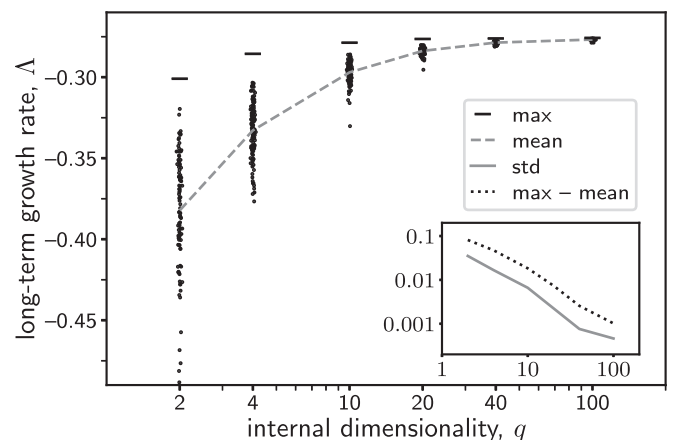

Fig. 5. Long-term population growth rate Λ vs. the dimensionality q of the intermediate layer of the network. Each point represents a network with a random, fixed representation matrix H (entries drawn from $\mathcal{N}(0, 1)$ independently) and an optimized expression matrix G. (To aid visualization of the density of points, a small random horizontal displacement is added.) Dashed and solid (*Inset*) lines show the mean and SD of the values of Λ. Horizontal bars mark the maximum values of Λ when the matrix H is also optimized for each q; dotted line (*Inset*) shows the difference between the maximum and the mean values. For this example the parameter values $\sigma = 1$ and $\gamma = 1$ are used.

To address this point, we consider the case where the representation matrix is chosen randomly. For a given dimensionality q, let each entry of H be drawn independently from a standard normal distribution $\mathcal{N}(0,1)$. For each of such random, fixed matrix H, the network is optimized over G to maximize the long-term population growth rate Λ. The results are shown in Fig. 5. We find that, for a relatively small q (such as $q = 4$), the values of Λ are low and widely spread; however, for a very large q (such as $q = 100$), the values of Λ are not only high but also narrowly distributed. Moreover, the distribution of Λ values moves closer to the maximum value as the dimensionality q increases. Hence, with a sufficiently high dimensionality, a random representation can be almost as good as the optimal one. This suggests that having a high-dimensional, sufficiently complex, internal representation of the environment would allow organisms to flexibly and quickly adapt to many situations. Of course, maintaining a large number of internal variables may incur additional costs.

The idea that a high-dimensional and potentially random representation of the input can encode complicated output patterns is related to the kernel method and reservoir computing in machine learning (23). In general, more complex patterns require higher dimensionality of the internal representation (see refs. 16 and 24 for discussion on the limitation of such methods). Similar ideas have been explored in biological contexts (19, 20).

Discussion

We have presented a general model of organisms' phenotypic responses to varying environments; the optimal responses show patterns of adaptation observed in nature. The form of such adaptation strategies depends on the noisiness of environmental cues and the selectivity of environmental conditions. In special limits of the parameter values, we have recovered three well-known strategies—unvarying, bet hedging, and tracking. The capacity of forming these and other adaptation strategies depends on the richness of the organisms' internal representation of the environment, characterized in our model by the number of internal variables.

Separation of Timescales. Our model implicitly assumes the separation of characteristic timescales of phenotypic responses, environmental changes, and evolution. In particular, by considering time in discrete numbers of generations, we do not model explicitly the dynamics of phenotypic development and environmental changes within a generation. This simplification is easily understood in cases where the timescale of environmental changes is much longer than that of the developmental process. In other cases, where the environment and the phenotype vary significantly within the lifetime, the vectors ε and ϕ can in principle represent time courses of the environment and the phenotype, respectively, such as growth conditions and behavioral traits during the lifetime of an organism. This would naturally make those vectors high dimensional and the mapping more complicated, which may inspire additional consideration on modeling the dynamics of phenotypic responses.

We have also assumed that the timescale of environmental changes is much shorter than that of evolutionary changes. This allowed us to consider the effect of evolution in varying environments by optimizing the environment-to-phenotype mapping with respect to the environmental statistics, without explicitly treating the dynamics of the evolutionary process. It should be noted that, when the timescale of environmental changes is comparable to that of evolutionary changes (such as the time for genetic mutations to arise and spread in a population), different modes of evolutionary dynamics may occur. Such situations have been theoretically studied in models of population genetics. For example, during a prolonged period of constant environment, organisms may lose the plasticity to express alternative phenotypes due to the accumulation of mutations affecting unused

phenotypes (25, 26). Similarly, bet hedging can be selected against in such a situation (27), and the population could go extinct before profiting from environmental changes.

When the environment is correlated over multiple generations, it is possible to reduce uncertainty in estimating the environment by tracking the history of environmental cues. This can be done by having organisms pass down information about their environment to their offspring, e.g., through epigenetic inheritance. Our current model does not include such a possibility, since the phenotype of an organism depends only on the environmental cue it receives and not on its parent's cue or phenotype. To incorporate transgenerational effects, one could, for example, let the state of the network in one generation depend on that in the previous generation, thus making the network recurrent across generations. Such generalization would allow the organisms to use temporal structures in the environmental variation.

Relation to Experiments. The geometry of phenotypic responses associated with different adaptation strategies can be looked for in experimental studies. Such studies should involve measuring the phenotype distribution in a wide range of controlled environmental conditions. Each strategy may be recognized by a particular shape of the phenotype distribution. For instance, an unvarying strategy is characterized by a phenotype distribution with a single peak that is stable under environmental variations. A pure bet-hedging strategy is associated with a multimodal phenotype distribution that does not depend on the environment. A tracking strategy, on the other hand, features a phenotype distribution with a single peak that changes position according to the environmental condition.

Our model predicts that specific adaptation strategies emerge under different levels of environmental noise and selection pressure. These predictions can be tested by experimental evolution. Indeed, several experiments have demonstrated that particular forms of adaptation can be evolved. For example, phenotypic plasticity, crucial for the tracking strategy in which organisms express distinctive phenotypes under varied environmental conditions, has been observed in larval development under temperature treatments (28). The evolution of bet-hedging strategies has been shown in bacteria subject to repeated selection in contrasting growth conditions (29). The random choice of phenotypes in a bet-hedging strategy may come from stochasticity in biochemical processes inside the organism. Alternatively, our model suggests that, when environmental cues are noisy and selection is strong, organisms can evolve to bet hedge using the cue as a source of randomness. Remarkably, a recent experiment in yeast showed that, indeed, bet hedging can be generated by plastic responses to an uninformative cue (30). Ultimately, a full test of our model requires varying the noise level of environmental cues and selection strength of environmental conditions and showing that different patterns of adaptation emerge from evolution. Such experiments would require quantitative and systematic measurements of the relation between organisms' phenotype and their environment.

Conclusion

We have introduced here the environment-to-phenotype mapping as an effective approach for studying the response of organisms to environmental conditions. This approach allowed us to explore a wide range of possible responses beyond the details of underlying molecular mechanisms. Compared with the commonly studied genotype-to-phenotype mapping, which describes how genetic variation affects phenotypes and emphasizes a mechanistic perspective (31–33), the environment-to-phenotype mapping provides a phenomenological perspective by describing organisms as a set of input–output relations that can be measured in experiments. This description is potentially useful for

studying evolution, since the same form of phenotypic responses may be naturally selected even if it is implemented by different molecular mechanisms. For instance, many bacteria can stochastically switch from a normal growth state to a dormant persister state, which prevents cell death from unforeseeable antibiotic attack (34). Different molecular mechanisms have been found to underlie such bacterial persistence (35). Nevertheless, the growth benefit of this particular adaptation strategy can be understood without using those mechanistic details (36). Such methods have recently been applied to other types of adaptation strategies (21, 37).

We have used a network model as a simple example of possible forms of the environment-to-phenotype mapping. In our model the connections of the network store information about the environmental conditions and their statistics, as well as about the favorable phenotypes. Besides varying the dimensionality of the internal representation or the number of intermediate layers (38), a possible further generalization of our model would be to consider a recurrent network with evolvable internal dynamics (39). Such a network could allow organisms to store information about their past phenotypes and encode temporal structures of the environmental history. The environment for the organisms can also include ecological interactions with individuals of the same population or other species. Such generalizations could lead to potentially more complex adaptation strategies.

Materials and Methods

Numerical Methods. Our goal is to maximize the long-term growth rate Λ with respect to the phenotypic response function Φ. The function Φ is parameterized by the matrices H and G, as in Eq. **3**. The value of Λ, according to Eq. **2**, is given by

$$\Lambda = \sum_\mu p_\mu \log F_\mu + \sum_\mu p_\mu \log \left\langle e^{-\frac{\gamma^2}{2}(\Phi(\xi)-\psi^\mu)^2} \right\rangle_{\xi \sim \mathcal{N}(\varepsilon^\mu, \sigma)} \quad \textbf{[5]}$$

where $\langle \cdot \rangle$ represents the expectation value with respect to the Gaussian random variable ξ. The first term does not depend on the parameters of Φ and is ignored. The optimization is done numerically by iterating over two steps: calculating the expectations in Eq. **5** given the current values of H and G, then updating these matrices to improve the value of Λ.

For the first step, we calculated the expectation values by numerically integrating over the Gaussian distributions. We used the python package "scipy.integrate," which calls the Fortran library QUADPACK. An alternative approach to numerical integration is to generate a random sample of ξ from the Gaussian distribution and use it to estimate the expectation values. This approach represents a finite sampling of the environmental cues, which allows for the analysis of the effect of finite population sizes and the stability of the optimal solutions. We tried both approaches and did not find significant differences in performance.

For the second step, we searched parameters using the python package "scipy.optimize" with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. This step involves calculating the gradient of the function Λ over the matrices H and G and then using the gradient to update their values. One could update the matrices simultaneously or optimize one while holding the other fixed and then iterate. It turns out that optimizing the matrix G alone is efficient, because G is directly connected to the output without having a nonlinear transformation. Using this observation, we chose to optimize G at every step of updating H. In this case, the gradient of $\Lambda(G^*(H), H)$ over H can be simply calculated as $\frac{\partial \Lambda}{\partial H}\Big|_{G^*}$ because $\frac{\partial \Lambda}{\partial G}\Big|_{G^*} = 0$.

For the examples shown in Fig. 3, the coordinates of the environments and the archetypes are $\varepsilon^1 = [-0.1, 0.9]$, $\varepsilon^2 = [-0.8, -0.4]$, $\varepsilon^3 = [0.9, -0.5]$, $\psi^1 = [-0.6, 0.5, 0.8]$, $\psi^2 = [0.4, 0.6, -0.9]$, $\psi^3 = [0.5, -0.8, 0.4]$; the environment probabilities are $[p_1, p_2, p_3] = [0.2, 0.5, 0.3]$. The same values are used for Figs. 4 and 5. In Fig. 4, for each pair of parameter values σ and γ, we ran eight replicate optimizations starting from random initial values (every entry of H and G drawn i.i.d. from $\mathcal{N}(0, 1)$); the order parameters are averaged over these replicates. In Fig. 5, for each dimensionality q, we ran 100 examples, each having a fixed H with random entries.

Analytic Limits. Nonparametrically, the ideal response function Φ^* that maximizes Eq. **5** should satisfy the variational equation $\delta\Lambda/\delta\Phi(\xi) = 0$,

which cannot be solved analytically. Here we derive approximate solutions for some extreme values of the parameters σ and γ. Our results in this subsection do not rely on the network ansatz, Eq. **3**, of the function Φ.

Weak selection, $\gamma \to 0$. In this limit, we can expand the integrand in Eq. **5**, to first order in γ^2, yielding

$$\Lambda \approx -\frac{\gamma^2}{2} \sum_\mu p_\mu \int d\xi\, P(\xi \mid \varepsilon^\mu)\, (\Phi(\xi) - \psi^\mu)^2, \quad \textbf{[6]}$$

where $P(\xi \mid \varepsilon^\mu)$ is the Gaussian distribution of ξ. To maximize the value of Λ, we set its variational derivative over the function $\Phi(\xi)$ to zero,

$$\frac{\delta\Lambda}{\delta\Phi(\xi)} = -\gamma^2 \sum_\mu p_\mu P(\xi \mid \varepsilon^\mu)\, (\Phi(\xi) - \psi^\mu) = 0. \quad \textbf{[7]}$$

Solving this equation yields

$$\Phi^*(\xi) = \frac{\sum_\mu p_\mu P(\xi \mid \varepsilon^\mu)\psi^\mu}{\sum_\nu p_\nu P(\xi \mid \varepsilon^\nu)} = \frac{\sum_\mu p_\mu \psi^\mu\, e^{-\frac{1}{2\sigma^2}(\xi - \varepsilon^\mu)^2}}{\sum_\nu p_\nu\, e^{-\frac{1}{2\sigma^2}(\xi - \varepsilon^\nu)^2}}. \quad \textbf{[8]}$$

This result can also be written succinctly as $\Phi^*(\xi) = \sum_\mu P(\varepsilon^\mu \mid \xi)\psi^\mu$, using Bayes' rule. The same expression has been derived in ref. 37.

In the subcase where σ is small, i.e., when the cue ξ is accurate, the probability $P(\varepsilon^\mu \mid \xi)$ is nearly 1 for the correct environment ε^μ, and hence the phenotypes are concentrated at the corresponding archetype ψ^μ. This yields the tracking strategy. However, when σ is large, i.e., when the cue is noisy, all environments ε^μ are likely; Eq. 8 becomes $\Phi^*(\xi) \approx \sum_\mu p_\mu \psi^\mu \equiv \bar{\psi}$, which means that an average phenotype $\bar{\psi}$ is produced regardless of the cue. This corresponds to the unvarying strategy.

Low noise, $\sigma \to 0$. In this limit, the Gaussian distribution of ξ in Eq. 5 is concentrated near its mean, ε^μ, so we can expand the integrand around that point. This yields, to first order in σ^2,

$$\Lambda \approx -\frac{\gamma^2}{2} \sum_\mu p_\mu \left[(\Phi(\varepsilon^\mu) - \psi^\mu)^2 + \sigma^2 \left(\partial_a \Phi_i(\varepsilon^\mu) \partial_a \Phi_i(\varepsilon^\mu) + \cdots \right) \right]. \quad \textbf{[9]}$$

This expression depends on the local values of the function Φ and its derivatives, $\Phi(\varepsilon^\mu)$, $\partial\Phi(\varepsilon^\mu)$, etc. To maximize Λ, we should have $\Phi^*(\varepsilon^\mu) \approx \psi^\mu$ and $\partial\Phi^*(\varepsilon^\mu) \approx 0$. It means that the ideal function Φ^* maps each environment ε^μ to its archetype ψ^μ, and the mapping is locally "flat"—the function value changes little in the neighborhood of ε^μ. Since, for low noise, the cues ξ are close to the actual environment ε^μ, they will all be mapped to near the correct archetype ψ^μ. This leads to the tracking strategy for any value of the selection strength γ.

High noise, $\sigma \to \infty$. In this limit, the cue ξ has a broad distribution that varies little with the environment ε^μ, and hence $P(\xi \mid \varepsilon^\mu) \approx P(\xi)$. As a result, the phenotype distribution will also be independent of the environment and can be defined as

$$\pi(\phi) \equiv \int d\xi\, P(\xi)\, \delta(\phi - \Phi(\xi)). \quad \textbf{[10]}$$

Using this phenotype distribution, the long-term growth rate Λ can be written as

$$\Lambda \approx \sum_\mu p_\mu \log \int d\phi\, \pi(\phi)\, e^{-\frac{\gamma^2}{2}(\phi - \psi^\mu)^2}. \quad \textbf{[11]}$$

The distribution $\pi^*(\phi)$ that maximizes Λ will constrain the ideal function Φ^* through Eq. **10**.

Let us treat the subcases of small and large γ separately. For a small γ, i.e., weak selection, we once again expand Λ to first order in γ^2, which yields

$$\Lambda \approx -\frac{\gamma^2}{2} \sum_\mu p_\mu \int d\phi\, \pi(\phi)\, (\phi - \psi^\mu)^2$$
$$= -\frac{\gamma^2}{2} \left(\int d\phi\, \pi(\phi)\, (\phi - \bar{\psi})^2 + \mathbb{V}[\psi] \right), \quad \textbf{[12]}$$

where $\mathbb{V}[\psi] = \sum_\mu p_\mu (\psi^\mu)^2 - \bar{\psi}^2$. From this expression it is clear that the optimal phenotype distribution is $\pi^*(\phi) = \delta(\phi - \bar{\psi})$, which agrees with the unvarying strategy found above.

For a large γ, it can be seen from Eq. 11 that the distribution $\pi(\phi)$ should become sharply peaked at points where $\phi = \psi^\mu$. We can use the ansatz
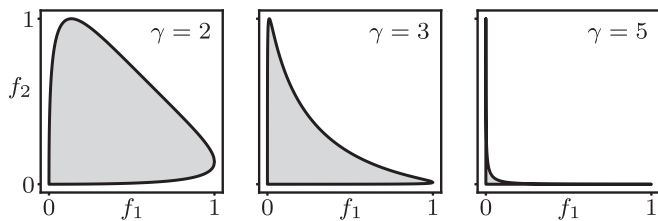
Fig. 6. Fitness sets (shaded area) for different values of the selection strength γ. Here f_1 and f_2 are fitness values of a phenotype in each of the two environments, with the corresponding archetypes separated by a distance $d = 1$.

$\pi(\phi) = \sum_\mu \pi_\mu \, \delta(\phi - \psi^\mu)$, which is a discrete distribution with weights only at the archetypes ψ^μ. Inserting this ansatz into Λ yields

$$\Lambda \approx \sum_\mu p_\mu \log \pi_\mu. \qquad [13]$$

This expression recovers the model of bet hedging (e.g., ref. 21). The optimal values of π_μ are given by $\pi_\mu^* = p_\mu$. Therefore, the phenotype distribution will consist of separate peaks at each ψ^μ, their relative sizes being proportional to the probability p_μ that each environment ε^μ occurs. To generate such a phenotype distribution, the function $\Phi^*(\xi)$ has to partition the environment space such that each partition has a total probability p_μ.

Strong selection, $\gamma \to \infty$. In this limit, the archetypes are far from one another as measured by the characteristic scale $1/\gamma$. Since a phenotype can be close to only one of the archetypes, there is a trade-off between the fitness values in different environments. In this case, the shape of the phenotype distribution can be understood by analyzing the geometry of the "fitness set" (8, 40).

Specifically, for each phenotype ϕ, the fitness values $f_\mu(\phi) \equiv f(\phi; \varepsilon^\mu)$ for $\mu = 1, \ldots, m$ can be represented by a point in an m-dimensional fitness

space. The collection of such points for all phenotypes ϕ forms the fitness set. Then, the average fitness of a population with a given phenotypic response function $\Phi(\xi)$ can be written as

$$f_\mu[\Phi] \equiv \int d\phi \, \pi(\phi \mid \varepsilon^\mu) f_\mu(\phi), \qquad [14]$$

where the phenotype distribution $\pi(\phi \mid \varepsilon^\mu)$ is given by

$$\pi(\phi \mid \varepsilon^\mu) \equiv \int d\xi \, P(\xi \mid \varepsilon^\mu) \, \delta(\phi - \Phi(\xi)). \qquad [15]$$

The collection of those points, $\{f_\mu[\Phi]\}$ for all possible phenotypic responses $\Phi(\xi)$, forms the "extended fitness set." Geometrically, each f_μ in the extended set can be considered as a linear combination of points from the original fitness set, weighted by the phenotype distribution in Eq. **14**. By locating the point within the extended fitness set that maximizes the long-term growth rate, $\Lambda = \sum_\mu p_\mu \log f_\mu$, one can find the optimal phenotypic response and the phenotype distribution (8).

As an example, consider two environments, $\mu = 1, 2$. The fitness values are given by $f_1 = e^{-\gamma^2(\phi - \psi^1)^2/2}$ and $f_2 = e^{-\gamma^2(\phi - \psi^2)^2/2}$, where the two archetypes are assumed to be at a distance $d = 1$ without loss of generality. In this case, the fitness set is shown in Fig. 6. It can be seen that, when $\gamma \gg 1$, the fitness set is highly concave. As a result, the extended fitness set will be largely formed by linear combinations of points near the corners at $(1, 0)$ and $(0, 1)$. This means that the phenotype distribution mainly consists of phenotypes near the archetypes ψ^1 and ψ^2. Hence, regardless of the cue, the optimal phenotype distribution will be peaked at the archetypes.

1. R. Kassen, The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002).
2. J. P. Sexton, J. Montiel, J. E. Shay, M. R. Stephens, R. A. Slatyer, Evolution of ecological niche breadth. *Annu. Rev. Ecol. Evol. Syst.* **48**, 183–206 (2017).
3. M. Slatkin, Hedging one's evolutionary bets. *Nature* **250**, 704–705 (1974).
4. A. M. Simons, Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. R. Soc. B* **278**, 1601–1609 (2011).
5. A. J. Grimbergen, J. Siebring, A. Solopova, O. P. Kuipers, Microbial bet-hedging: The power of being different. *Curr. Opin. Microbiol.* **25**, 67–72 (2015).
6. T. J. DeWitt, A. Sih, D. S. Wilson, Costs and limits of phenotypic plasticity. *Trends Ecol. Evol.* **13**, 77–81 (1998).
7. A. P. Hendry, Key questions on the role of phenotypic plasticity in eco-evolutionary dynamics. *J. Hered.* **107**, 25–41 (2016).
8. A. Mayer, T. Mora, O. Rivoire, A. M. Walczak, Transitions in optimal adaptive strategies for populations in fluctuating environments. *Phys. Rev. E* **96**, 032412 (2017).
9. P. R. Grant, *Ecology and Evolution of Darwin's Finches* (Princeton University Press, 1986).
10. O. Shoval et al., Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336**, 1157–1160 (2012).
11. A. P. Moczek et al., The role of developmental plasticity in evolutionary innovation. *Proc. R. Soc. B* **278**, 2705–2713 (2011).
12. D. Cohen, Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12**, 119–129 (1966).
13. D. L. Venable, Bet hedging in a guild of desert annuals. *Ecology* **88**, 1086–1090 (2007).
14. D. Cohen, Optimizing reproduction in a randomly varying environment when a correlation may exist between the conditions at the time a choice has to be made and the subsequent outcome. *J. Theor. Biol.* **16**, 1–14 (1967).
15. J. R. Gremer, S. Kimball, D. L. Venable, Within- and among-year germination in Sonoran desert winter annuals: Bet hedging and predictive germination in a variable environment. *Ecol. Lett.* **19**, 1209–1218 (2016).
16. J. Hertz, R. G. Palmer, A. S. Krogh, *Introduction to the Theory of Neural Computation* (Perseus Publishing, ed. 1, 1991).
17. B. Scheres, W. H. van der Putten, The plant perceptron connects environment to development. *Nature* **543**, 337–345 (2017).
18. W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, B. Goldstein, The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* **84**, 783–794 (2003).
19. B. Babadi, H. Sompolinsky, Sparseness and expansion in sensory representations. *Neuron* **83**, 1213–1226 (2014).
20. K. Krishnamurthy, A. M. Hermundstad, T. Mora, A. M. Walczak, V. Balasubramanian, Disorder and the neural representation of complex odors: Smelling in the real world. arXiv:1707.01962 (6 July 2017).
21. O. Rivoire, S. Leibler, The value of information for populations in varying environments. *J. Stat. Phys.* **142**, 1124–1166 (2011).
22. M. C. Donaldson-Matasci, C. T. Bergstrom, M. Lachmann, When unreliable cues are good enough. *Am. Nat.* **182**, 313–327 (2013).
23. M. Lukosevicius, H. Jaeger, Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**, 127–149 (2009).
24. M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning* (MIT Press, 2012).
25. U. Gerland, T. Hwa, Evolutionary selection between alternative modes of gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8841–8846 (2009).
26. J. Masel, O. D. King, H. Maughan, The loss of adaptive plasticity during long periods of environmental stasis. *Am. Nat.* **169**, 38–46 (2007).
27. O. D. King, J. Masel, The evolution of bet-hedging adaptations to rare scenarios. *Theor. Popul. Biol.* **72**, 560–575 (2007).
28. Y. Suzuki, H. F. Nijhout, Evolution of a polyphenism by genetic accommodation. *Science* **311**, 650–652 (2006).
29. H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, P. B. Rainey, Experimental evolution of bet hedging. *Nature* **462**, 90–93 (2009).
30. C. S. Maxwell, P. M. Magwene, When sensing is gambling: An experimental system reveals how plasticity can generate tunable bet-hedging strategies. *Evolution* **71**, 859–871 (2017).
31. P. Alberch, From genes to phenotype: Dynamical systems and evolvability. *Genetica* **84**, 5–11 (1991).
32. M. Pigliucci, Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philos. Trans. R. Soc. B* **365**, 557–566 (2010).
33. S. E. Ahnert, Structural properties of genotype-phenotype maps. *J. R. Soc. Interf.* **14**, 20170275 (2017).
34. N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
35. N. R. Cohen, M. A. Lobritz, J. J. Collins, Microbial persistence and the road to drug resistance. *Cell Host Microbe* **13**, 632–642 (2013).
36. E. Kussell, S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
37. B. K. Xue, S. Leibler, Benefits of phenotypic plasticity for population growth in varying environments. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12745–12750 (2018).
38. T. Friedlander, A. E. Mayo, T. Tlusty, U. Alon, Evolution of bow-tie architectures in biology. *PLoS Comput. Biol.* **11**, e1004055 (2015).
39. K. Kaneko, Evolution of robustness and plasticity under environmental fluctuation: Formulation in terms of phenotypic variances. *J. Stat. Phys.* **148**, 687–705 (2012).
40. R. Levins, Evolution in changing environments: Some theoretical explorations. *Monographs in Population Biology* (Princeton University Press, 1968).